

WHAT IS CLAIMED IS:

1. A method of analyzing a computer program, the method comprising:

running code of the computer program over a plurality of intervals of execution;

5 during said step of executing code, tracking a statistic for a program component;

identifying a behavior of the computer program over each of the plurality of intervals of execution based on the tracked statistic;

10 comparing at least one identified behavior for at least one interval of execution to another interval of execution to determine similarity between the intervals of execution.

2. The method of claim 1 wherein said step of running code comprises at least one of executing the program on hardware, simulating the program's execution in software, direct execution, emulating the program's execution in software, and modeling a hypothetical execution in software.

5

3. The method of claim 2 wherein the statistic comprises at least one of a hardware metric and a hardware-independent metric.

4. The method of claim 3 wherein the statistic comprises at least one of frequency of the component occurring in execution, number of

instructions executed, amount of memory used by the program component, time, IPC, performance counters, program counters, and cache miss rate.

5

5. The method of claim 2 wherein the program component comprises an identifiable section of control flow of the computer program.

6. The method of claim 5 wherein the program component comprises at least one of an instruction, basic block, procedure, loop, load instruction, and branch instruction.

7. The method of claim 2 wherein the program component comprises a memory region.

8. The method of claim 5 wherein the program component comprises a basic block, the basic block being a section of the code having a single entry point and a single exit.

9. The method of claim 2 wherein each of the plurality of intervals of execution comprises at least one of a time interval, an instruction interval, and a metric-based interval.

10. The method of claim 9 wherein one of the plurality of intervals comprises the execution of at least one of the subset of the code and the full execution of the code.

11. The method of claim 9 wherein the intervals of execution comprise at least one of overlapping and non-overlapping intervals.

12. The method of claim 2 further comprising:  
based on said comparing step, classifying the plurality of intervals of execution into at least one cluster, wherein each of the intervals is more likely to be similar in program behavior to the other intervals in that  
5 cluster than to the intervals in a remainder of clusters.

13. The method of claim 12 further comprising:  
selecting at least one representative interval of execution for each  
of the at least one cluster.

14. The method of claim 13 wherein each of the at least one representative interval of execution is closest to an average behavior of the cluster.

15. The method of claim 13 wherein the representative interval of execution is the earliest interval of execution within a predetermined distance from an average behavior of the cluster.

16. The method of claim 13 further comprising:

weighing each of the selected representative intervals of execution based on at least one of a total amount of time, a number of instructions within the cluster, the program component, and the statistic.

5

17. The method of claim 16 wherein the weighted representative intervals collectively represent a complete execution of at least a subset of the computer program.

18. The method of claim 12 further comprising:

minimizing the number of clusters.

19. The method of claim 17 further comprising:

minimizing the number of clusters.

20. The method of claim 14 further comprising:

weighing each of the selected representative intervals of execution based on at least one of a total amount of time, a number of instructions within the cluster, the program component, and the statistic;

5                    wherein the weighted representative intervals collectively  
represent a complete execution of at least a subset of the computer program;  
and further comprising minimizing the number of clusters.

21.    The method of claim 15 further comprising:  
weighing each of the selected representative intervals of  
execution based on at least one of a total amount of time, a number of  
instructions within the cluster, the program component, and the statistic;

5                    wherein the weighted representative intervals collectively  
represent a complete execution of at least a subset of the computer program  
and further comprising minimizing the number of clusters.

22.    The method of claim 10 wherein the step of comparing  
comprises:  
comparing each interval to the interval of execution representing  
at least a subset of execution of the computer program.

5                    23.    The method of claim 22 further comprising:  
based on said comparing step, identifying an end of an  
initialization of the computer program.

24.    The method of claim 22 further comprising:

based on said comparing step, identifying a length of at least one repeating interval of execution.

25. The method of claim 24 wherein said step of identifying a length comprises performing an analysis of a signal, the signal comprising differences between each identified interval of execution and the interval of execution representing the at least a subset of execution of the computer  
5 program.

26. The method of claim 12 further comprising:  
determining a confidence and variance by sampling the intervals of execution within a particular cluster for at least one of a hardware metric and a hardware-independent metric.

5

27. A method of analyzing a computer program, the method comprising:  
running at least a portion of the computer program;  
identifying behavior of a hardware-independent metric within at  
5 least one arbitrary section of execution of the portion of the computer program during said executing step;  
classifying each of the at least one arbitrary section of execution according to the identified behavior into clusters of similar behavior.

28. The method of claim 27 wherein said step of identifying comprises identifying a frequency of execution of basic blocks of the executed code, wherein each of the at least one basic block comprises a piece of code of the computer program executed from start to finish, said basic block having only one entry point and one exit.

29. The method of claim 28 wherein said step of identifying a frequency provides a group of frequencies for each of the number of intervals.

30. The method of claim 29 wherein said step of classifying further comprises:

comparing the identified behavior of one of the intervals to the identified behavior of another of the intervals to identify a phase of the interval.

5

31. The method of claim 33 further comprising:

identifying an initialization phase;

determining at least one analysis point occurring after execution of the identified initialization phase.

5

32. The method of claim 28 wherein said step of identifying a frequency of execution for each of the at least one basic block comprises:

for each of the number of intervals, determining a interval vector,  
the interval vector comprising a plurality of ordered elements, each of the  
5 plurality of ordered elements relating to a particular basic block and  
representing a frequency of execution of the particular basic block.

33. The method of claim 28 further comprising:  
partitioning the computer program into a set of clusters by  
comparing the determined interval vectors to one another.

34. The method of claim 33 wherein said step of partitioning  
further comprises:

determining a group of clusters;  
comparing each of the interval vectors to each of the set of  
5 clusters;  
adding the compared interval vector to a particular cluster based  
on a goodness of fit between the compared basic block vectors and each of the  
group of clusters;  
changing a centroid of each of the group of clusters;  
10 repeating the comparing, adding, and clustering steps to form the  
set of clusters.

35. A method of analyzing operation of a computer program,  
the method comprising:



executing at least a portion of the computer program;  
for each of a plurality of intervals of execution over the at least a  
5 portion of the computer program, identifying behavior of a hardware-  
independent metric;  
identifying behavior of the hardware-independent metric over full  
execution of the at least a portion of the computer program to identify a target  
behavior;  
10 comparing the identified behavior of each of the plurality of  
intervals to the identified target behavior over full execution to determine a  
representative interval;  
simulating execution of the computer program over the  
determined representative interval.

15

36. The method of claim 35 wherein said step of identifying  
comprises:

deriving a plurality of basic block vectors, each basic block  
vector representing code blocks of the program executed during the interval of  
5 execution, the basic block vector being based on frequencies of basic blocks of  
executed code within execution of the program;

wherein the basic block vector comprises a single dimensional  
array where a single element in the array exists for a basic block in the  
program.

10

37. The method of claim 2 wherein the method is performed in run-time.

38. The method of claim 37 wherein said step of identifying comprises tracking a proportion of instructions executed from different sections of code of the program over each of the plurality of intervals;

further comprising, for each interval, classifying the identified  
5 behavior into phases corresponding to changes in behavior across the executed code.

39. The method of claim 38 further comprising:  
predicting when execution of the code is about to enter a phase  
change;

predicting a phase entered by the phase change.

40. The method of claim 38 wherein said step of identifying comprises, for each section of code:

capturing an identifier of the section of code;

capturing a number of instructions executed for the section of  
5 code.

41. The method of claim 40 further comprising:

reducing a number of the identified sections of code to a lower number of buckets.

42. The method of claim 40 further comprising:

comparing each section of code to a previous section of code in a history;

if the compared section of code is not similar to the previous  
5 section of code, adding the previous section of code to the history.

43. The method of claim 2 wherein the behavior identified for an interval is collected in a vector, the vector containing the statistic for at least one element representing at least one component.

44. The method of claim 43 wherein the vector is retained in at least one of a memory, a storage medium, and a table.

45. The method of claim 44 wherein the vector is stored as a signature that represents at least one of the behavior of a complete vector, a projection of the vector, a compressed representation of the vector, a partial representation of the vector, and a subset of the behavior identified collected in  
5 the vector.

46. The method of claim 45 further comprising:

storing a phase ID with the signature, wherein the phase ID comprises at least one of a complete signature, a subset of the signature, a partial representation of the signature, a name independent of the signature, and  
5 a number.

47. The method of claim 46 wherein the stored phase ID is identified for an interval by at least one of looking up the signature in storage, and if an approximate match exists, using the phase ID stored with the signature, and creating a new phase ID.

5

48. The method of claim 47 wherein the identified behavior and the tracked statistic for at least one interval with a phase ID are stored and associated with one another.

49. The method of claim 48 wherein, if a storage area for storing the phase ID, behavior, and statistic is finite, only a single stored signature for a phase ID, and the phase ID, are stored.

50. The method of claim 47 further comprising, after the phase ID is identified by a signature for an interval:

looking up the phase ID to find the associated statistic for the interval.

5

51. The method of claim 50 further comprising:

using the found associated statistic, performing at least one of a behavior optimization, statistic optimization, load-time optimization, run-time optimization, and hardware reconfiguration.

5

52. The method of claim 47 wherein the phase ID is stored in a prediction table, and further comprising:

predicting a phase ID for an interval using the stored phase ID.

53. The method of claim 52 further comprising:

retrieving information for the predicted phase ID;

using the retrieved information, guiding optimization for the computer program.

5